Chapter 9
Martingales

One of the most important concepts of modern probability theory is the martingale, which formalizes the notion of a fair game. In this chapter, we first lay the foundations for the treatment of general stochastic processes. We then introduce martingales and the discrete stochastic integral. We close with an application to a model from mathematical finance.

9.1 Processes, Filtrations, Stopping Times

We introduce the fundamental technical terms for the investigation of stochastic processes (including martingales). In order to be able to recycle the terms later in a more general context, we go for greater generality than is necessary for the treatment of martingales only.

In the following, let (E, τ) be a Polish space with Borel σ-algebra $\mathcal{E}$. Further, let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and let $I \subset \mathbb{R}$ be arbitrary. We are mostly interested in the cases $I = \mathbb{N}_0$, $I = \mathbb{Z}$, $I = [0, \infty)$ and I an interval.

Definition 9.1 (Stochastic process) Let $I \subset \mathbb{R}$. A family of random variables $X = (X_t, t \in I)$ (on $(\Omega, \mathcal{F}, \mathbf{P})$) with values in $(E, \mathcal{E})$ is called a *stochastic process* with index set (or time set) I and range E.

Remark 9.2 Sometimes families of random variables with more general index sets are called stochastic processes. We come back to this with the Poisson point process in Chapter 24. $\Diamond$

Remark 9.3 Following a certain tradition, we will often denote a stochastic process by $X = (X_t)_{t \in I}$ if we want to emphasize the "time evolution" aspect rather than the formal notion of a family of random variables. Formally, both objects are of course the same. $\Diamond$

A. Klenke, *Probability Theory*, Universitext, DOI 10.1007/978-1-4471-5361-0_9,

Example 9.4 Let $I = \mathbb{N}_0$ and let $(Y_n, n \in \mathbb{N})$ be a family of i.i.d. $\mathrm{Rad}_{1/2}$-random variables on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$; that is, random variables with

$$\mathbf{P}[Y_n = 1] = \mathbf{P}[Y_n = -1] = \frac{1}{2}.$$

Let $E = \mathbb{Z}$ (with the discrete topology) and let

$$X_t = \sum_{n=1}^{t} Y_n \quad \text{for all } t \in \mathbb{N}_0.$$

$(X_t, t \in \mathbb{N}_0)$ is called a *symmetric simple random walk* on $\mathbb{Z}$. ◊

Example 9.5 The Poisson process $X = (X_t)_{t \geq 0}$ with intensity $\alpha > 0$ (see Section 5.5) is a stochastic process with range $\mathbb{N}_0$. ◊

We introduce some further terms.

Definition 9.6 If X is a random variable (or a stochastic process), we write $\mathcal{L}[X] = \mathbf{P}_X$ for the distribution of X. If $\mathcal{G} \subset \mathcal{F}$ is a σ-algebra, then we write $\mathcal{L}[X \mid \mathcal{G}]$ for the regular conditional distribution of X given $\mathcal{G}$.

Definition 9.7 An E-valued stochastic process $X = (X_t)_{t \in I}$ is called

(i) *real-valued* if $E = \mathbb{R}$,
(ii) a process with *independent increments* if X is real-valued and for all $n \in \mathbb{N}$ and all $t_0, \ldots, t_n \in I$ with $t_0 < t_1 < \ldots < t_n$, we have that

$$(X_{t_i} - X_{t_{i-1}})_{i=1,\ldots,n} \quad \text{is independent,}$$

(iii) a *Gaussian process* if X is real-valued and for all $n \in \mathbb{N}$ and $t_1, \ldots, t_n \in I$,

$$(X_{t_1}, \ldots, X_{t_n}) \quad \text{is } n\text{-dimensional normally distributed, and}$$

(iv) *integrable* (respectively *square integrable*) if X is real-valued and $\mathbf{E}[|X_t|] < \infty$ (respectively $\mathbf{E}[(X_t)^2] < \infty$) for all $t \in I$.
Now assume that $I \subset \mathbb{R}$ is closed under addition. Then X is called
(v) *stationary* if $\mathcal{L}[(X_{s+t})_{t \in I}] = \mathcal{L}[(X_t)_{t \in I}]$ for all $s \in I$, and
(vi) a process with *stationary increments* if X is real-valued and

$$\mathcal{L}[X_{s+t+r} - X_{t+r}] = \mathcal{L}[X_{s+r} - X_r] \quad \text{for all } r, s, t \in I.$$

(If $0 \in I$, then it is enough to consider $r = 0$.)

Example 9.8

(i) The Poisson process with intensity θ and the random walk on $\mathbb{Z}$ are processes with stationary independent increments.

(ii) If X_t, $t \in I$, are i.i.d. random variables, then $(X_t)_{t \in I}$ is stationary.
(iii) Let $(X_n)_{n \in \mathbb{Z}}$ be real-valued and stationary and let $k \in \mathbb{N}$ and $c_0, \ldots, c_k \in \mathbb{R}$. Define

$$Y_n := \sum_{i=0}^{k} c_i X_{n-i}.$$

Then $Y = (Y_n)_{n \in \mathbb{Z}}$ is a stationary process. If $c_0, \ldots, c_k \geq 0$ and $c_0 + \ldots + c_k = 1$, then Y is called the *moving average* of X (with weights $c_0, \ldots, c_k$). ◊

The following two definitions make sense also for more general index sets I that are partially ordered. However, we restrict ourselves to the case $I \subset \mathbb{R}$.

Definition 9.9 (Filtration) Let $\mathbb{F} = (\mathcal{F}_t, t \in I)$ be a family of σ-algebras with $\mathcal{F}_t \subset \mathcal{F}$ for all $t \in I$. $\mathbb{F}$ is called a *filtration* if $\mathcal{F}_s \subset \mathcal{F}_t$ for all $s, t \in I$ with $s \leq t$.

Definition 9.10 A stochastic process $X = (X_t, t \in I)$ is called *adapted* to the filtration $\mathbb{F}$ if X_t is $\mathcal{F}_t$-measurable for all $t \in I$. If $\mathcal{F}_t = \sigma(X_s, s \leq t)$ for all $t \in I$, then we denote by $\mathbb{F} = \sigma(X)$ the filtration that is generated by X.

Remark 9.11 Clearly, a stochastic process is always adapted to the filtration it generates. Hence the generated filtration is the smallest filtration to which the process is adapted. ◊

Definition 9.12 (Predictable) A stochastic process $X = (X_n, n \in \mathbb{N}_0)$ is called *predictable* (or *previsible*) with respect to the filtration $\mathbb{F} = (\mathcal{F}_n, n \in \mathbb{N}_0)$ if X_0 is constant and if, for every $n \in \mathbb{N}$

$$X_n \text{ is } \mathcal{F}_{n-1}\text{-measurable.}$$

Example 9.13 Let $I = \mathbb{N}_0$ and let $Y_1, Y_2, \ldots$ be real random variables. For $n \in \mathbb{N}_0$, define $X_n := \sum_{m=1}^{n} Y_m$. Let

$$\mathcal{F}_0 = \{\emptyset, \Omega\} \quad \text{and} \quad \mathcal{F}_n = \sigma(Y_1, \ldots, Y_n) \quad \text{for } n \in \mathbb{N}.$$

Then $\mathbb{F} = (\mathcal{F}_n, n \in \mathbb{N}_0) = \sigma(Y)$ is the filtration generated by $Y = (Y_n)_{n \in \mathbb{N}}$ and X is adapted to $\mathbb{F}$; hence $\sigma(X) \subset \mathbb{F}$. Clearly, $(Y_1, \ldots, Y_n)$ is measurable with respect to $\sigma(X_1, \ldots, X_n)$; hence $\sigma(Y) \subset \sigma(X)$, and thus also $\mathbb{F} = \sigma(X)$.

Now let $\widetilde{X}_n := \sum_{m=1}^{n} \mathbb{1}_{[0,\infty)}(Y_m)$. Then $\widetilde{X}$ is also adapted to $\mathbb{F}$; however, in general, $\mathbb{F} \supsetneq \sigma(\widetilde{X})$. ◊

Example 9.14 Let $I = \mathbb{N}_0$ and let $D_1, D_2, \ldots$ be i.i.d. $\text{Rad}_{1/2}$-distributed random variables (that is, $\mathbf{P}[D_i = -1] = \mathbf{P}[D_i = 1] = \frac{1}{2}$ for all $i \in \mathbb{N}$). Let $D = (D_i)_{i \in \mathbb{N}}$ and $\mathbb{F} = \sigma(D)$. We interpret D_i as the result of a bet that gives a gain or loss of one euro for every euro we put at stake. Just before each gamble we decide how much money we bet. Let H_n be the number of euros to bet in the nth gamble. Clearly, H_n may only depend on the results of the gambles that happened earlier, but not on D_m for any $m \geq n$. To put it differently, there must be a function $F_n : \{-1, 1\}^{n-1} \to \mathbb{N}$ such that $H_n = F_n(D_1, \ldots, D_{n-1})$. (For example, for the Petersburg game (Example 4.22) we had $F_n(x_1, \ldots, x_{n-1}) = 2^{n-1}\mathbb{1}_{\{x_1=x_2=\ldots=x_{n-1}=0\}}$.) Hence H is predictable. On the other hand, any predictable H has the form $H_n = F_n(D_1, \ldots, D_{n-1})$, $n \in \mathbb{N}$, for certain functions $F_n : \{-1, 1\}^{n-1} \to \mathbb{N}$. Hence any predictable H is an admissible gambling strategy. $\Diamond$

Definition 9.15 (Stopping time) A random variable τ with values in $I \cup \{\infty\}$ is called a *stopping time* (with respect to $\mathbb{F}$) if for any $t \in I$

$$\{\tau \leq t\} \in \mathcal{F}_t.$$

The idea is that $\mathcal{F}_t$ reflects the knowledge of an observer at time t. Whether or not $\{\tau \leq t\}$ is true can thus be determined on the basis of the information available at time t.

Theorem 9.16 *Let I be countable. τ is a stopping time if and only if $\{\tau = t\} \in \mathcal{F}_t$ for all $t \in I$.*

Proof This is left as an exercise! $\square$

Example 9.17 Let $I = \mathbb{N}_0$ (or, more generally, let $I \subset [0, \infty)$ be right-discrete; that is, $t < \inf I \cap (t, \infty)$ for all $t \geq 0$, and hence I in particular is countable) and let $K \subset \mathbb{R}$ be measurable. Let X be an adapted real-valued stochastic process. Consider the first time that X is in K:

$$\tau_K := \inf\{t \in I : X_t \in K\}.$$

It is intuitively clear that τ_K should be a stopping time since we can determine by observation up to time t whether $\{\tau_K \leq t\}$ occurs. Formally, we argue that $\{X_s \in K\} \in \mathcal{F}_s \subset \mathcal{F}_t$ for all $s \leq t$. Hence also the countable union of these sets is in $\mathcal{F}_t$:

$$\{\tau_K \leq t\} = \bigcup_{s \in I \cap [0,t]} \{X_s \in K\} \in \mathcal{F}_t.$$

Consider now the random time $\widetilde{\tau} := \sup\{t \in I : X_t \in K\}$ of the last visit of X to K. For a fixed time t, on the basis of previous observations, we cannot determine whether X is already in K for the last time. For this we would have to rely on prophecy. Hence, in general, $\widetilde{\tau}$ is not a stopping time. $\Diamond$

Lemma 9.18 *Let $I \subset [0, \infty)$ be closed under addition and let σ and τ be stopping times. Then:*

(i) *$\sigma \vee \tau$ and $\sigma \wedge \tau$ are stopping times.*
(ii) *If $\sigma, \tau \geq 0$, then $\sigma + \tau$ is also a stopping time.*
(iii) *If $s \geq 0$, then $\tau + s$ is a stopping time. However, in general, $\tau - s$ is not.*

Before we present the (simple) formal proof, we state that in particular (i) and (iii) are properties we would expect of stopping times. With (i), the interpretation is clear. For (iii), note that $\tau - s$ peeks into the future by s time units (in fact, $\{\tau - s \leq t\} \in \mathcal{F}_{t+s}$), while $\tau + s$ looks back s time units. For stopping times, however, only retrospection is allowed.

Proof (i) For $t \in I$, we have $\{\sigma \vee \tau \leq t\} = \{\sigma \leq t\} \cap \{\tau \leq t\} \in \mathcal{F}_t$ and $\{\sigma \wedge \tau \leq t\} = \{\sigma \leq t\} \cup \{\tau \leq t\} \in \mathcal{F}_t$.

(ii) Let $t \in I$. By (i), $\tau \wedge t$ and $\sigma \wedge t$ are stopping times for any $t \in I$. In particular, $\{\tau \wedge t \leq s\} \in \mathcal{F}_s \subset \mathcal{F}_t$ for any $s \leq t$. On the other hand, we have $\tau \wedge t \leq s$ for $s > t$. Hence $\tau' := (\tau \wedge t) + \mathbb{1}_{\{\tau > t\}}$ and $\sigma' := (\sigma \wedge t) + \mathbb{1}_{\{\sigma > t\}}$ (and thus $\tau' + \sigma'$) are $\mathcal{F}_t$-measurable. We conclude $\{\tau + \sigma \leq t\} = \{\tau' + \sigma' \leq t\} \in \mathcal{F}_t$.

(iii) For $\tau + s$, this is a consequence of (ii) (with the stopping time $\sigma \equiv s$). For $\tau - s$, since τ is a stopping time, we have $\{\tau - s \leq t\} = \{\tau \leq t + s\} \in \mathcal{F}_{t+s}$. However, in general, $\mathcal{F}_{t+s}$ is a strict superset of $\mathcal{F}_t$; hence $\tau - s$ is not a stopping time. $\square$

Definition 9.19 Let τ be a stopping time. Then

$$\mathcal{F}_\tau := \left\{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F}_t \text{ for any } t \in I\right\}$$

is called the *σ-algebra of τ-past.*

Example 9.20 Let $I = \mathbb{N}_0$ (or let $I \subset [0, \infty)$ be right-discrete; compare Example 9.17) and let X be an adapted real-valued stochastic process. Let $K \in \mathbb{R}$ and let $\tau = \inf\{t : X_t \geq K\}$ be the stopping time of first entrance in $[K, \infty)$. Consider the events $A = \{\sup\{X_t : t \in I\} > K - 5\}$ and $B = \{\sup\{X_t : t \in I\} > K + 5\}$.

Clearly, $\{\tau \leq t\} \subset A$ for all $t \in I$; hence $A \cap \{\tau \leq t\} = \{\tau \leq t\} \in \mathcal{F}_t$. Thus $A \in \mathcal{F}_\tau$. However, in general, $B \notin \mathcal{F}_\tau$ since up to time τ, we cannot decide whether X will ever exceed $K + 5$. $\Diamond$

Lemma 9.21 *If σ and τ are stopping times with $\sigma \leq \tau$, then $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$.*

Proof Let $A \in \mathcal{F}_\sigma$ and $t \in I$. Then $A \cap \{\sigma \leq t\} \in \mathcal{F}_t$. Now $\{\tau \leq t\} \in \mathcal{F}_t$ since τ is a stopping time. Since $\sigma \leq \tau$, we thus get

$$A \cap \{\tau \leq t\} = \big(A \cap \{\sigma \leq t\}\big) \cap \{\tau \leq t\} \in \mathcal{F}_t. \qquad \square$$

Definition 9.22 If $\tau < \infty$ is a stopping time, then we define $X_\tau(\omega) := X_{\tau(\omega)}(\omega)$.

Lemma 9.23 *Let I be countable, let X be adapted and let $\tau < \infty$ be a stopping time. Then X_τ is measurable with respect to $\mathcal{F}_\tau$.*

Proof Let A be measurable and $t \in I$. Hence $\{\tau = s\} \cap X_s^{-1}(A) \in \mathcal{F}_s \subset \mathcal{F}_t$ for all $s \le t$. Thus

$$X_\tau^{-1}(A) \cap \{\tau \le t\} = \bigcup_{\substack{s\in I\\ s\le t}} \big(\{\tau = s\} \cap X_s^{-1}(A)\big) \in \mathcal{F}_t.$$

□

For uncountable I and for fixed ω, in general, the map $I \to E$, $t \mapsto X_t(\omega)$ is not measurable; hence neither is the composition X_τ always measurable. Here one needs assumptions on the regularity of the *paths* $t \mapsto X_t(\omega)$; for example, right continuity. We come back to this point in Chapter 21 and leave this as a warning for the time being.

9.2 Martingales

Everyone who does not own a casino would agree without hesitation that the successive payment of gains $Y_1, Y_2, \ldots$, such that $Y_1, Y_2, \ldots$ are i.i.d. with $\mathbf{E}[Y_1] = 0$, could be considered a fair game consisting of consecutive rounds. In this case, the process X of partial sums $X_n = Y_1 + \ldots + Y_n$ is integrable and $\mathbf{E}[X_n \mid \mathcal{F}_m] = X_m$ if $m < n$ (where $\mathbb{F} = \sigma(X)$). We want to use this equation for the conditional expectations as the defining equation for a fair game that in the following will be called a martingale. Note that, in particular, this definition does not require that the individual payments be independent or identically distributed. This makes the notion quite a bit more flexible. The momentousness of the following concept will become manifest only gradually.

Definition 9.24 Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, $I \subset \mathbb{R}$, and let $\mathbb{F}$ be a filtration. Let $X = (X_t)_{t\in I}$ be a real-valued, adapted stochastic process with $\mathbf{E}[|X_t|] < \infty$ for all $t \in I$. X is called (with respect to $\mathbb{F}$) a

martingale if $\mathbf{E}[X_t \mid \mathcal{F}_s] = X_s$ for all $s, t \in I$ with $t > s$,
submartingale if $\mathbf{E}[X_t \mid \mathcal{F}_s] \ge X_s$ for all $s, t \in I$ with $t > s$,
supermartingale if $\mathbf{E}[X_t \mid \mathcal{F}_s] \le X_s$ for all $s, t \in I$ with $t > s$.

Remark 9.25 Clearly, for a martingale, the map $t \mapsto \mathbf{E}[X_t]$ is constant, for submartingales it is monotone increasing and for supermartingales it is monotone decreasing. ◊

Remark 9.26 The etymology of the term *martingale* has not been resolved completely. The French *la martingale* (originally Provençal *martegalo*, named after the town *Martiques*) in equitation means “a piece of rein used in jumping and cross country riding". Sometimes the ramified shape, in particular of the *running martingale* (French *la martingale à anneaux*), is considered as emblematic for the doubling strategy in the Petersburg game.

This doubling strategy itself is the second meaning of *la martingale*. Starting here, a shift in the meaning towards the mathematical notion seems plausible. A different derivation, in contrast to the appearance, is based on the function of the rein, which is to "check the upward movement of the horse's head". Thus the notion of a martingale might first have been used for general gambling strategies (checking the movements of chance) and later for the doubling strategy in particular. ◊

Remark 9.27 If $I = \mathbb{N}$, $I = \mathbb{N}_0$ or $I = \mathbb{Z}$, then it is enough to consider at each instant s only $t = s + 1$. In fact, by the tower property of the conditional expectation (Theorem 8.14(iv)), we get

$$\mathbf{E}[X_{s+2} \mid \mathcal{F}_s] = \mathbf{E}\big[\mathbf{E}[X_{s+2} \mid \mathcal{F}_{s+1}] \,\big|\, \mathcal{F}_s\big].$$

Thus, if the defining equality (or inequality) holds for any time step of size one, by induction it holds for all times. ◊

Remark 9.28 If we do not explicitly mention the filtration $\mathbb{F}$, we tacitly assume that $\mathbb{F}$ is generated by X; that is, $\mathcal{F}_t = \sigma(X_s, s \le t)$. ◊

Remark 9.29 Let $\mathbb{F}$ and $\mathbb{F}'$ be filtrations with $\mathcal{F}_t \subset \mathcal{F}'_t$ for all t, and let X be an $\mathbb{F}'$-(sub-, super-) martingale that is adapted to $\mathbb{F}$. Then X is also a (sub-, super-) martingale with respect to the smaller filtration $\mathbb{F}$. Indeed, for $s < t$ and for the case of a submartingale,

$$\mathbf{E}[X_t \mid \mathcal{F}_s] = \mathbf{E}\big[\mathbf{E}\big[X_t \,\big|\, \mathcal{F}'_s\big] \,\big|\, \mathcal{F}_s\big] \ge \mathbf{E}[X_s \mid \mathcal{F}_s] = X_s.$$

In particular, an $\mathbb{F}$-(sub-, super-) martingale X is always a (sub-, super-) martingale with respect to its own filtration $\sigma(X)$. ◊

Example 9.30 Let $Y_1, \ldots, Y_N$ be independent random variables with $\mathbf{E}[Y_t] = 0$ for all $t = 1, \ldots, N$. Let $\mathcal{F}_t := \sigma(Y_1, \ldots, Y_t)$ and $X_t := \sum_{s=1}^{t} Y_s$. Then X is adapted and integrable, and $\mathbf{E}[Y_r \mid \mathcal{F}_s] = 0$ for $r > s$. Hence, for $t > s$,

$$\mathbf{E}[X_t \mid \mathcal{F}_s] = \mathbf{E}[X_s \mid \mathcal{F}_s] + \mathbf{E}[X_t - X_s \mid \mathcal{F}_s] = X_s + \sum_{r=s+1}^{t} \mathbf{E}[Y_r \mid \mathcal{F}_s] = X_s.$$

Thus, X is an $\mathbb{F}$-martingale.

Similarly, X is a submartingale if $\mathbf{E}[Y_t] \ge 0$ for all t, and a supermartingale if $\mathbf{E}[Y_t] \le 0$ for all t. ◊

Example 9.31 Consider the situation of the preceding example; however, now with $\mathbf{E}[Y_t] = 1$ and $X_t = \prod_{s=1}^{t} Y_s$ for $t \in \mathbb{N}_0$. By Theorem 5.4, $Y_1 \cdot Y_2$ is integrable.

Inductively, we get $\mathbf{E}[|X_t|] < \infty$ for all $t \in \mathbb{N}_0$. Evidently, X is adapted to $\mathbb{F}$ and for all $s \in \mathbb{N}_0$, we have

$$\mathbf{E}[X_{s+1} \mid \mathcal{F}_s] = \mathbf{E}[X_s Y_{s+1} \mid \mathcal{F}_s] = X_s \mathbf{E}[Y_{s+1} \mid \mathcal{F}_s] = X_s.$$

Hence X is an $\mathbb{F}$-martingale. $\Diamond$

Theorem 9.32

(i) *X is a supermartingale if and only if $(-X)$ is a submartingale.*
(ii) *Let X and Y be martingales and let $a, b \in \mathbb{R}$. Then $(aX + bY)$ is a martingale.*
(iii) *Let X and Y be supermartingales and $a, b \geq 0$. Then $(aX + bY)$ is a supermartingale.*
(iv) *Let X and Y be supermartingales. Then $Z := X \wedge Y = (\min(X_t, Y_t))_{t \in I}$ is a supermartingale.*
(v) *If $(X_t)_{t \in \mathbb{N}_0}$ is a supermartingale and $\mathbf{E}[X_T] \geq \mathbf{E}[X_0]$ for some $T \in \mathbb{N}_0$, then $(X_t)_{t \in \{0,\ldots,T\}}$ is a martingale. If there exists a sequence $T_N \to \infty$ with $\mathbf{E}[X_{T_N}] \geq \mathbf{E}[X_0]$, then X is a martingale.*

Proof (i), (ii) and (iii) These are evident.

(iv) Since $|Z_t| \leq |X_t| + |Y_t|$, we have $\mathbf{E}[|Z_t|] < \infty$ for all $t \in I$. Due to monotonicity of the conditional expectation (Theorem 8.14(ii)), for $t > s$, we have $\mathbf{E}[Z_t \mid \mathcal{F}_s] \leq \mathbf{E}[Y_t \mid \mathcal{F}_s] \leq Y_s$ and $\mathbf{E}[Z_t \mid \mathcal{F}_s] \leq \mathbf{E}[X_t \mid \mathcal{F}_s] \leq X_s$; hence

$$\mathbf{E}[Z_t \mid \mathcal{F}_s] \leq X_s \wedge Y_s = Z_s.$$

(v) For $t \leq T$, let $Y_t := \mathbf{E}[X_T \mid \mathcal{F}_t]$. Then Y is a martingale and $Y_t \leq X_t$. Hence

$$\mathbf{E}[X_0] \leq \mathbf{E}[X_T] = \mathbf{E}[Y_T] = \mathbf{E}[Y_t] \leq \mathbf{E}[X_t] \leq \mathbf{E}[X_0].$$

(The first inequality holds by assumption.) We infer that $Y_t = X_t$ almost surely for all t and thus $(X_t)_{t \in \{0,\ldots,T\}}$ is a martingale.

Let $T_N \to \infty$ with $\mathbf{E}[X_{T_N}] \geq \mathbf{E}[X_0]$ for all $N \in \mathbb{N}$. Then, for any $t > s \geq 0$, there is an $N \in \mathbb{N}$ with $T_N > t$. Hence, $\mathbf{E}[X_t \mid \mathcal{F}_s] = \mathbf{E}[X_s]$ and X is a martingale. $\square$

Remark 9.33 Many statements about supermartingales hold *mutatis mutandis* for submartingales. For example, in the preceding theorem, claim (i) holds with the words “submartingale” and “supermartingale” interchanged, claim (iv) holds for submartingales if the minimum is replaced by a maximum, and so on. We often do not give the statements both for submartingales and for supermartingales. Instead, we choose representatively one case. Note, however, that those statements that we make explicitly about martingales usually cannot be adapted easily to sub- or supermartingales (such as (ii) in the preceding theorem). $\Diamond$

Corollary 9.34 *Let X be a submartingale and $a \in \mathbb{R}$. Then $(X - a)^+$ is a submartingale.*

Proof Clearly, 0 and $Y = X - a$ are submartingales. By (iv), $(X - a)^+ = Y \vee 0$ is also a submartingale. □

Theorem 9.35 *Let X be a martingale and let $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function.*

(i) *If*

$$\mathbf{E}\big[\varphi(X_t)^+\big] < \infty \quad \text{for all } t \in I, \tag{9.1}$$

then $(\varphi(X_t))_{t\in I}$ is a submartingale.

(ii) *If $t^* := \sup(I) \in I$, then $\mathbf{E}[\varphi(X_{t^*})^+] < \infty$ implies (9.1).*

(iii) *In particular, if $p \geq 1$ and $\mathbf{E}[|X_t|^p] < \infty$ for all $t \in I$, then $(|X_t|^p)_{t\in I}$ is a submartingale.*

Proof (i) We always have $\mathbf{E}[\varphi(X_t)^-] < \infty$ (Theorem 7.9); hence, by assumption, $\mathbf{E}[|\varphi(X_t)|] < \infty$ for all $t \in I$. Jensen's inequality (Theorem 8.20) then yields, for $t > s$,

$$\mathbf{E}\big[\varphi(X_t) \,\big|\, \mathcal{F}_s\big] \geq \varphi\big(\mathbf{E}[X_t \mid \mathcal{F}_s]\big) = \varphi(X_s).$$

(ii) Since φ is convex, so is $x \mapsto \varphi(x)^+$. Furthermore, by assumption, we have $\mathbf{E}[\varphi(X_{t^*})^+] < \infty$; hence Jensen's inequality implies that, for all $t \in I$,

$$\begin{aligned}\mathbf{E}\big[\varphi(X_t)^+\big] &= \mathbf{E}\Big[\varphi\big(\mathbf{E}[X_{t^*} \mid \mathcal{F}_t]\big)^+\Big] \leq \mathbf{E}\Big[\mathbf{E}\big[\varphi(X_{t^*})^+ \,\big|\, \mathcal{F}_t\big]\Big] \\ &= \mathbf{E}\big[\varphi(X_{t^*})^+\big] < \infty.\end{aligned}$$

(iii) This is evident since $x \mapsto |x|^p$ is convex. □

Example 9.36 (See Example 9.4.) Symmetric simple random walk X on $\mathbb{Z}$ is a square integrable martingale. Hence $(X_n^2)_{n\in\mathbb{N}_0}$ is a submartingale. ◊

Exercise 9.2.1 Let Y be a random variable with $\mathbf{E}[|Y|] < \infty$ and let $\mathbb{F}$ be a filtration as well as

$$X_t := \mathbf{E}[Y \mid \mathcal{F}_t] \quad \text{for all } t \in I.$$

Show that X is an $\mathbb{F}$-martingale.

Exercise 9.2.2 Let $(X_n)_{n\in\mathbb{N}_0}$ be a predictable $\mathbb{F}$-martingale. Show that $X_n = X_0$ almost surely for all $n \in \mathbb{N}_0$.

Exercise 9.2.3 Show that the claim of Theorem 9.35 continues to hold if X is only a submartingale but if φ is in addition assumed to be monotone increasing. Give an example that shows that the monotonicity of φ is essential. (Compare Corollary 9.34.)

Exercise 9.2.4 (Azuma's inequality) Show the following.

(i) If X is a random variable with $|X| \le 1$ a.s., then there is a random variable Y with values in $\{-1, +1\}$ and with $\mathbf{E}[Y \mid X] = X$.

(ii) For X as in (i) with $\mathbf{E}[X] = 0$, infer that (using Jensen's inequality)

$$\mathbf{E}\big[e^{\lambda X}\big] \le \cosh(\lambda) \le e^{\lambda^2/2} \quad \text{for all } \lambda \in \mathbb{R}.$$

(iii) If $(M_n)_{n\in\mathbb{N}_0}$ is a martingale with $M_0 = 0$ and if there is a sequence $(c_k)_{k\in\mathbb{N}}$ of nonnegative numbers with $|M_n - M_{n-1}| \le c_n$ a.s. for all $n \in \mathbb{N}$, then

$$\mathbf{E}\big[e^{\lambda M_n}\big] \le \exp\left(\frac{1}{2}\lambda^2 \sum_{k=1}^{n} c_k^2\right).$$

(iv) Under the assumptions of (iii), *Azuma's inequality* holds:

$$\mathbf{P}\big[|M_n| \ge \lambda\big] \le 2\exp\left(-\frac{\lambda^2}{2\sum_{k=1}^{n} c_k^2}\right) \quad \text{for all } \lambda \ge 0.$$

Hint: Use Markov's inequality for $f(x) = e^{\gamma x}$ and choose the optimal γ.

9.3 Discrete Stochastic Integral

So far we have encountered a martingale as the process of partial sums of gains of a fair game. This game can also be the price of a stock that is traded at discrete times on a stock exchange. With this interpretation, it is particularly evident that it is natural to construct new stochastic processes by considering investment strategies for the stock. The value of the portfolio, which is the new stochastic process, changes as the stock price changes. It is the price multiplied by the number of stocks in the portfolio. In order to describe such processes formally, we introduce the following notion.

Definition 9.37 (Discrete stochastic integral) Let $(X_n)_{n\in\mathbb{N}_0}$ be an $\mathbb{F}$-adapted real process and let $(H_n)_{n\in\mathbb{N}}$ be a real-valued and $\mathbb{F}$-predictable process. The discrete stochastic integral of H with respect to X is the stochastic process $H\cdot X$ defined by

$$(H\cdot X)_n := \sum_{m=1}^{n} H_m (X_m - X_{m-1}) \quad \text{for } n \in \mathbb{N}_0. \tag{9.2}$$

If X is a martingale, then $H\cdot X$ is also called the *martingale transform* of X.

Remark 9.38 Clearly, $H\cdot X$ is adapted to $\mathbb{F}$. ◊

Let X be a (possibly unfair) game where $X_n - X_{n-1}$ is the gain per euro in the nth round. We interpret H_n as the number of euros we bet in the nth game. H is then a *gambling strategy*. Clearly, the value of H_n has to be decided at time $n-1$; that is, *before* the result of X_n is known. In other words, H must be predictable.

Now assume that X is a fair game (that is, a martingale) and H is *locally bounded* (that is, each H_n is bounded). Then (since $\mathbf{E}[X_{n+1} - X_n \mid \mathcal{F}_n] = 0$)

$$\begin{aligned}\mathbf{E}\big[(H\cdot X)_{n+1} \,\big|\, \mathcal{F}_n\big] &= \mathbf{E}\big[(H\cdot X)_n + H_{n+1}(X_{n+1} - X_n) \,\big|\, \mathcal{F}_n\big] \\ &= (H\cdot X)_n + H_{n+1}\mathbf{E}[X_{n+1} - X_n \mid \mathcal{F}_n] \\ &= (H\cdot X)_n.\end{aligned}$$

Thus $H\cdot X$ is a martingale. The following theorem says that the converse also holds; that is, X is a martingale if, for sufficiently many predictable processes, the stochastic integral is a martingale.

Theorem 9.39 (Stability theorem) *Let $(X_n)_{n\in\mathbb{N}_0}$ be an adapted, real-valued stochastic process with $\mathbf{E}[|X_0|] < \infty$.*

(i) *X is a martingale if and only if, for any locally bounded predictable process H, the stochastic integral $H\cdot X$ is a martingale.*
(ii) *X is a submartingale (supermartingale) if and only if $H\cdot X$ is a submartingale (supermartingale) for any locally bounded predictable $H \geq 0$.*

Proof (i) "$\Longrightarrow$" This has been shown in the discussion above.

"$\Longleftarrow$" Fix an $n_0 \in \mathbb{N}$, and let $H_n = \mathbb{1}_{\{n=n_0\}}$. Then $(H\cdot X)_{n_0-1} = 0$; hence

$$0 = \mathbf{E}\big[(H\cdot X)_{n_0} \,\big|\, \mathcal{F}_{n_0-1}\big] = \mathbf{E}[X_{n_0} \mid \mathcal{F}_{n_0-1}] - X_{n_0-1}.$$

(ii) This is similar to (i). $\square$

The preceding theorem says, in particular, that we cannot find any locally bounded gambling strategy that transforms a martingale (or, if we are bound to nonnegative gambling strategies, as we are in real life, a supermartingale) into a submartingale. Quite the contrary is suggested by the many invitations to play all kinds of "sure winning systems" in lotteries.

Example 9.40 (Petersburg game) We continue Example 9.14 (see also Example 4.22). Define $X_n := D_1 + \ldots + D_n$ for $n \in \mathbb{N}_0$. Then X is a martingale. The gambling strategy $H_n := 2^{n-1}\mathbb{1}_{\{D_1=D_2=\ldots=D_{n-1}=-1\}}$ for $n \in \mathbb{N}$ and $H_0 = 1$ is predictable and locally bounded. Let $S_n = \sum_{i=1}^n H_i D_i = (H\cdot X)_n$ be the gain after n rounds. Then S is a martingale by the preceding theorem. In particular, we get (as shown already in Example 4.22) that $\mathbf{E}[S_n] = 0$ for all $n \in \mathbb{N}$. We will come back later to the point that this superficially contrasts with $S_n \overset{n\to\infty}{\longrightarrow} 1$ a.s. (see Example 11.6).

For the moment, note that the martingale $S' = (1 - S_n)_{n\in\mathbb{N}_0}$, just like the one in Example 9.31, has the structure of a product of independent random variables with expectation 1. In fact, $S'_n = \prod_{i=1}^n (1 - D_i)$. $\diamond$

9.4 Discrete Martingale Representation Theorem and the CRR Model

By virtue of the stochastic integral, we have transformed a martingale X via a gambling strategy H into a new martingale $H\cdot X$. Let us change the perspective and ask: For fixed X, which are the martingales Y (with $Y_0 = 0$) that can be obtained as discrete stochastic integrals of X with a suitable gambling strategy $H = H(Y)$? Possibly all martingales Y? This is not the case, in general, as the example below indicates. However, we will see that all martingales can be represented as stochastic integrals if the increments $X_{n+1} - X_n$ can take only two values (given $X_1, \dots, X_n$). In this case, we give a representation theorem and use it to discuss the fair price for a European call option in the stock market model of Cox–Ross–Rubinstein. This model is rather simple and describes an idealized market (no transaction costs, fractional numbers of stocks tradeable and so on). For extensive literature on stochastic aspects of mathematical finance, we refer to the textbooks [9, 42, 48, 57, 86, 102, 121] or [159].

Example 9.41 Consider the very simple martingale $X = (X_n)_{n=0,1}$ with only two time points. Let $X_0 = 0$ almost surely and $\mathbf{P}[X_1 = -1] = \mathbf{P}[X_1 = 0] = \mathbf{P}[X_1 = 1] = \frac{1}{3}$. Let $Y_0 = 0$. Further, let $Y_1 = 2$ if $X_1 = 1$ and $Y_1 = -1$ otherwise. Then Y is manifestly a $\sigma(X)$-martingale. However, there is no number H_1 such that $H_1 X_1 = Y_1$. $\Diamond$

Let $T \in \mathbb{N}$ be a fixed time. If $(Y_n)_{n=0,1,\dots,T}$ is an $\mathbb{F}$-martingale, then $Y_n = \mathbf{E}[Y_T \mid \mathcal{F}_n]$ for all $n \le T$. An $\mathbb{F}$-martingale Y is thus determined uniquely by the terminal values Y_T (and vice versa). Let X be a martingale. As $(H\cdot X)$ is a martingale, the representation problem for martingales is thus reduced to the problem of representing an integrable random variable $V := Y_T$ as $v_0 + (H\cdot X)_T$, where $v_0 = \mathbf{E}[Y_T]$.

We saw that, in general, this is not possible if the differences $X_{n+1} - X_n$ take three (or more) different values. Hence we now consider the case where only two values are possible. Here, at each time step, a system of two linear equations with two unknowns has to be solved. In the case where $X_{n+1} - X_n$ takes three values, the system has three equations and is thus overdetermined.

Definition 9.42 (Binary model) A stochastic process $X_0, \dots, X_T$ is called *binary splitting* or a *binary model* if there exist random variables $D_1, \dots, D_T$ with values in $\{-1, +1\}$ and functions $f_n : \mathbb{R}^{n-1} \times \{-1, +1\} \to \mathbb{R}$ for $n = 1, \dots, T$, as well as $x_0 \in \mathbb{R}$ such that $X_0 = x_0$ and

$$X_n = f_n(X_1, \dots, X_{n-1}, D_n) \quad \text{for any } n = 1, \dots, T.$$

By $\mathbb{F} = \sigma(X)$, we denote the filtration generated by X.

Note that X_n depends only on $X_1, \dots, X_{n-1}$ and D_n but not on the full information inherent in the values $D_1, \dots, D_n$. If the latter were the case, a ramification into more than two values in one time step would be possible.

Theorem 9.43 (Representation theorem) *Let X be a binary model and let V_T be an $\mathcal{F}_T$-measurable random variable. Then there exists a bounded predictable process H and a $v_0 \in \mathbb{R}$ with $V_T = v_0 + (H \cdot X)_T$.*

Note that $\mathbb{F}$ is the filtration generated by X, not the, possibly larger, filtration generated by $D_1, \dots, D_T$. For the latter case, the claim of the theorem would be incorrect since, loosely speaking, with H we can bet on X but not on the D_i.

Proof We show that there exist $\mathcal{F}_{T-1}$-measurable random variables V_{T-1} and H_T such that $V_T = V_{T-1} + H_T(X_T - X_{T-1})$. By a backward induction, this yields the claim.

Since V_T is $\mathcal{F}_T$-measurable, by the factorization lemma (Corollary 1.97) there exists a function $g_T : \mathbb{R}^T \to \mathbb{R}$ with $V_T = g_T(X_1, \dots, X_T)$. Define

$$X_T^{\pm} = f_T(X_1, \dots, X_{T-1}, \pm 1) \quad \text{and} \quad V_T^{\pm} = g_T\big(X_1, \dots, X_{T-1}, X_T^{\pm}\big).$$

Each of these four random variables is manifestly $\mathcal{F}_{T-1}$-measurable. Hence we are looking for solutions V_{T-1} and H_T of the following system of linear equations:

$$\begin{aligned} V_{T-1} + H_T\big(X_T^- - X_{T-1}\big) &= V_T^-, \\ V_{T-1} + H_T\big(X_T^+ - X_{T-1}\big) &= V_T^+. \end{aligned} \tag{9.3}$$

By construction, $X_T^+ - X_T^- \neq 0$ if $V_T^+ - V_T^- \neq 0$. Hence we can solve (9.3) and get

$$H_T := \begin{cases} \frac{V_T^+ - V_T^-}{X_T^+ - X_T^-}, & \text{if } X_T^+ \neq X_T^-, \\ 0, & \text{else,} \end{cases}$$

and $V_{T-1} = V_T^+ - H_T(X_T^+ - X_{T-1}) = V_T^- - H_T(X_T^- - X_{T-1})$. □

We now want to interpret X as the market price of a stock and V_T as the payment of a financial derivative on X, a so-called *contingent claim* or, briefly, claim. For example, V_T could be a (European) *call option* with *maturity* T and *strike price* $K \geq 0$. In this case, we have $V_T = (X_T - K)^+$. Economically speaking, the European call gives the buyer the right (but not the obligation) to buy one stock at time T at price K (from the issuer of the option). As typically the option is exercised only if the market price at time T is larger than K (and then gives a profit of $X_T - K$ as the stock could be sold at price X_T on the market), the value of the option is indeed $V_T = (X_T - K)^+$.

At the stock exchanges, not only are stocks traded but also derivatives on stocks. Hence, what is the fair price $\pi(V_T)$ for which a trader would offer (and buy) the contingent claim V_T? If there exists a strategy H and a v_0 such that $V_T = v_0 + (H \cdot X)_T$, then the trader can sell the claim for v_0 (at time 0) and replicate the claim by building a portfolio that follows the trading strategy H. In this case, the claim V_T is called *replicable* and the strategy H is called a *hedging strategy*, or briefly a hedge. A market in which every claim can be replicated is called complete. In this sense, the binary model is a complete market.

If there was a second strategy H' and a second v_0' with $v_0' + (H' \cdot X)_T = V_T$, then, in particular, $v_0 - v_0' = ((H' - H) \cdot X)_T$. If we had $v_0 > v_0'$, then the trader could follow the strategy $H' - H$ (which gives a final payment of $V_T - V_T = 0$) and make a sure profit of $v_0 - v_0'$. In the opposite case, $v_0 < v_0'$, the strategy $H - H'$ ensures a risk-free profit. Such a risk-free profit (or *free lunch* in economic jargon) is called an *arbitrage*. It is reasonable to assume that a market gives no opportunity for an arbitrage. Hence the fair price $\pi(V_T)$ is determined uniquely once there is *one* trading strategy H and a v_0 such that $V_T = v_0 + (H \cdot X)_T$.

Until now, we have not assumed that X is a martingale. However, if X is a martingale, then $(H \cdot X)$ is a martingale with $(H \cdot X)_0 = 0$; hence clearly $\mathbf{E}[(H \cdot X)_T] = 0$. Thus

$$\pi(V_T) = v_0 = \mathbf{E}[V_T]. \tag{9.4}$$

Since, in this case, v_0 does not depend on the trading strategy and is hence unique, the market is automatically arbitrage-free. A finite market is thus arbitrage-free if and only if there exists an equivalent martingale (to be defined below). In this case, uniqueness of this martingale is equivalent to completeness of the market ("the fundamental theorem of asset pricing" by Harrison–Pliska [68]). In larger markets, equivalence holds only with a somewhat more flexible notion of arbitrage (see [30]).

Now if X is not a martingale, then in some cases, we can replace X by a different process X' that *is* a martingale and such that the distributions $\mathbf{P}_X$ and $\mathbf{P}_{X'}$ are equivalent; that is, have the same null sets. Such a process is called an *equivalent martingale*, and $\mathbf{P}_{X'}$ is called an *equivalent martingale measure*. A trading strategy that replicates V_T with respect to X also replicates V_T with respect to X'. In particular, the fair price does not change if we pass to the equivalent martingale X'. Thus we can compute $\pi(V_T)$ by applying (9.4) to the equivalent martingale.

While here this is only of interest in that it simplifies the computation of fair prices, it has an economic interpretation as a measure for the market prices that we would see if all traders were risk-neutral; that is, for traders who price a future payment by its mean value. Typically, however, traders are risk-averse and thus real market prices include a discount due to the inherent risk.

Now we consider one model in greater detail.

Definition 9.44 Let $T \in \mathbb{N}$, $a \in (-1, 0)$ and $b > 0$ as well as $p \in (0, 1)$. Further, let $D_1, \ldots, D_T$ be i.i.d. Rad_p random variables (that is, $\mathbf{P}[D_1 = 1] = 1 - \mathbf{P}[D_1 = -1] = p$). We let $X_0 = x_0 > 0$ and for $n = 1, \ldots, T$, define

$$X_n = \begin{cases} (1+b)X_{n-1}, & \text{if } D_n = +1, \\ (1+a)X_{n-1}, & \text{if } D_n = -1. \end{cases}$$

X is called the *multi-period binomial model* or the *Cox–Ross–Rubinstein model* (without interest returns).

As we have shown already, the CRR model is complete. Further, with the choice $p = p^* := \frac{a}{a-b}$, we can change X into a martingale. Hence the model is also arbitrage-free (for all $p \in (0, 1)$). Now we want to compute the price of a European call option $V_T := (X_T - K)^+$ explicitly. To this end, we can assume $p = p^*$. Letting $A := \min\{i \in \mathbb{N}_0 : (1+b)^i (1+a)^{T-i} x_0 > K\}$, we get

$$\begin{aligned}
\pi(V_T) = \mathbf{E}_{p^*}[V_T] &= \sum_{i=0}^{T} b_{T,p^*}\big(\{i\}\big)\big[(1+b)^i (1+a)^{T-i} x_0 - K\big]^+ \\
&= x_0 \sum_{i=A}^{T} \binom{T}{i} \big(p^*\big)^i \big(1-p^*\big)^{T-i} \big[(1+b)^i (1+a)^{T-i}\big] \\
&\quad - K \sum_{i=A}^{T} b_{T,p^*}\big(\{i\}\big).
\end{aligned}$$

If we define $p' = (1+b)p^*$, then $p' \in (0, 1)$ and $1 - p' = (1-p^*)(1+a)$. We thus obtain the Cox–Ross–Rubinstein formula

$$\pi(V_T) = x_0 b_{T,p'}\big(\{A, \ldots, T\}\big) - K b_{T,p^*}\big(\{A, \ldots, T\}\big). \tag{9.5}$$

This is the discrete analogue of the celebrated Black–Scholes formula for option pricing in certain time-continuous markets.